

THE k -DIMENSIONAL DISTRIBUTION OF COMBINED GFSR SEQUENCES

SHU TEZUKA

ABSTRACT. We develop an efficient method for analysis of the k -dimensional distribution of combinations of several GFSR sequences by bitwise exclusive-or operations. First, we introduce the notion of a resolution-wise lattice structure for GFSR sequences, and show that by applying a theorem of Couture to this type of lattice, we obtain a precise description of k -dimensional distribution of combined GFSR sequences in the same way as for combined Tausworthe sequences. Finally, we apply this method to the combination of two different Twisted GFSR generators, which were recently proposed by Matsumoto and Kurita, and investigate the order of equidistribution of the combined sequence.

1. INTRODUCTION

Couture et al. [1] have recently developed an efficient method to give a precise description of how all the k -dimensional vectors formed by successive values of simple or combined Tausworthe sequences are distributed in the unit hypercube, based on their lattice structure in the space of formal Laurent series over $GF(2)$. This method can be applied to a special subclass of GFSR sequences as well as their combinations with bitwise exclusive-or operations (XORs) if they can be formulated as linear congruential sequences in the field of formal Laurent series over $GF(2)$, but it cannot be applied to general GFSR sequences such as Twisted GFSR generators proposed by Matsumoto and Kurita [6]. These generators, a subclass of GFSR sequences, have the attractive property that they can generate very long-period GFSR sequences, using a minimum amount of memory, almost as fast as the conventional GFSR algorithm, but it has been found that all simple Twisted GFSRs have a deficiency of uniform distribution properties in dimensions higher than the degree of the recurrence relation. For this reason, the XOR-combination of Twisted GFSRs has been investigated as a possible way to overcome this defect. As stated above, however, the problem is that we lack an efficient method for investigating combined Twisted GFSRs.

The objective of this paper is to develop a theoretical tool for analysis of the k -dimensional distribution of XOR-combinations of general GFSRs on the basis of Couture's theorem [1]. The paper is organized as follows: §2 overviews the definition of combined GFSR sequences and Couture et al.'s results for

Received by the editor January 19, 1993.

1991 *Mathematics Subject Classification.* Primary 65C10.

Key words and phrases. GFSR sequences, Twisted GFSR, k -dimensional distribution, lattice structure.

the lattice structure of combined Tausworthe sequences. In §3, we develop a method for analysis of the k -dimensional distribution of combined GFSR sequences. First, we introduce the notion of a resolution-wise lattice for GFSR sequences, and show that we can apply Couture's theorem to this type of lattice, thereby obtaining a precise description of the k -dimensional distribution of GFSR sequences in the same way as for combined Tausworthe sequences. In §4, we give an example in which our approach is applied to a combined Twisted GFSR in practical use with a period length of about 2^{1200} . Section 5 discusses the efficiency of our approach.

2. OVERVIEW

2.1. GFSR sequences. An (L -bit) GFSR sequence u_i , $i = 1, 2, \dots$, is originally defined as follows [5]:

$$(1) \quad u_i = \sum_{j=1}^L b_{dj+i} 2^{-j},$$

where b_j , $j = 1, 2, \dots$, is a linear feedback shift register sequence modulo two whose characteristic polynomial $M(z)$ is primitive over F_2 . Lewis and Payne [5] suggested that d should be greater than $100r$, where $r = \deg(M)$. In addition, they employed a primitive trinomial as the characteristic polynomial of the binary sequence b_j , $j = 1, 2, \dots$, in order to realize a fast generation scheme for the sequence in the following way: Let $M(z) = z^r + z^s + 1$ ($r > s$). The sequence can then be generated by the scheme

$$u_i = u_{i-r+s} \text{ XOR } u_{i-r}.$$

However, this algorithm is unsatisfactory in the sense that the period $2^r - 1$ is much smaller than the maximum possible period $2^{Lr} - 1$ attainable by using r L -bit words.

A more general version of the sequence is defined, for $i = 1, 2, \dots$, as

$$(2) \quad u_i = \sum_{l=1}^L b_{j_l+i} 2^{-l},$$

where j_l , $l = 1, \dots, L$, are integers between 0 and $2^r - 1$, and the characteristic polynomial $M(z)$ is any primitive polynomial [3, 8, 9]. As shown in [12], Tausworthe sequences can be viewed as a subclass of GFSR sequences.

The matrix representation of shift register sequences is very useful. Let C be the companion matrix of the polynomial $M(z) = z^r + a_{r-1}z^{r-1} + \dots + a_1z + a_0$, namely,

$$C = \begin{pmatrix} 0 & 0 & 0 & \cdots & a_0 \\ 1 & 0 & 0 & \cdots & a_1 \\ & \cdots & \cdots & \cdots & \\ & \cdots & \cdots & \cdots & \\ 0 & \cdots & 1 & 0 & a_{r-2} \\ 0 & 0 & \cdots & 1 & a_{r-1} \end{pmatrix},$$

and let $\mathbf{b} = (b_0, \dots, b_{r-1})$ be a nonzero binary vector. The general GFSR sequence is then written as follows:

$$\mathbf{b}G, \mathbf{b}CG, \dots, \mathbf{b}C^iG, \dots,$$

where G is an $r \times L$ matrix over F_2 whose l th column vector, denoted by $G_l, l = 1, \dots, L$, is uniquely determined by the equations,

$$b_{j+i} = (G_l, \mathbf{b}C^{i-1}) \text{ for } i = 1, \dots, r.$$

Here (\mathbf{a}, \mathbf{b}) means the inner product of the binary vectors \mathbf{a} and \mathbf{b} over F_2 .

Now, we give some definitions relevant to the k -dimensional distribution of GFSR sequences. Let $S(l, k) = \{\alpha^{jh+i}, 1 \leq i \leq k, 1 \leq h \leq l (\leq L)\}$, where α is the root of $M(z)$.

Definition 1. A GFSR sequence is said to be k -distributed with l -bit resolution if $S(l, k)$ is linearly independent, but $S(l+1, k)$ is not, for some $l \leq L$.

The following is also useful:

Definition 2. A GFSR sequence is said to have the order of equidistribution k for the leading l bits if $S(l, k)$ is linearly independent, but $S(l, k+1)$ is not.

Twisted GFSR generators were recently defined by Matsumoto and Kurita [6]:

$$(3) \quad \mathbf{v}_i = \mathbf{v}_{i-r+s} \text{ XOR } \mathbf{v}_{i-r}A \text{ for } i = r+1, r+2, \dots,$$

where $\mathbf{v}_i, i = 1, 2, \dots$, is a sequence of vectors in F_2^L, A is an $L \times L$ matrix over F_2 , and $r > s$. The parameters r, s , and A are chosen so that the maximum period of the sequence $\mathbf{v}_i, i = 1, 2, \dots$, becomes $2^{Lr} - 1$. In particular, Matsumoto and Kurita analyzed the following special case, which is very useful for quick generation of the sequences:

$$(4) \quad \mathbf{v}_i = \mathbf{v}_{i-r+s} \text{ XOR } \mathbf{v}_{i-r}C^T,$$

where $\mathbf{v}_i, i = 1, 2, \dots$, is a sequence of vectors in F_2^L, C is an $L \times L$ companion matrix, and $r > s$. The conversion from a binary vector $\mathbf{v} = (v_0, \dots, v_{L-1})$ to a random number between 0 and 1 is as follows:

$$u = \frac{1}{2^L} \sum_{i=0}^{L-1} v_i 2^{L-1-i}.$$

Their paper lists some parameters of the Twisted GFSR in (4) with maximum periods. Note that Twisted GFSRs with maximum period lengths $2^{Lr} - 1$ can be formulated as in (2) with $b_j, j = 1, 2, \dots$, being a shift register sequence of maximum period $2^{Lr} - 1$.

One of the advantages of this scheme is that the size r of an array is the minimum necessary to produce GFSR sequences with the period length $2^{Lr} - 1$ with respect to the wordsize L . In this sense, the scheme can be viewed as an improved version of the lagged Fibonacci scheme with XOR, because the latter produces a period $2^r - 1$ of a sequence with an identically sized array of r L -bit words. Another important merit is the fast generation algorithm, which

is as follows: Let $\mathbf{v}_i = (v_{i,0}, \dots, v_{i,L-1})$ and $\mathbf{a} = (a_0, \dots, a_{L-1})$. The rest is very simple:

$$\begin{aligned} \text{if } v_{i-r,L-1} = 0 \text{ then } \mathbf{v}_i &= \mathbf{v}_{i-r+s} \text{ XOR SR}(\mathbf{v}_{i-r}) \\ \text{else } \mathbf{v}_i &= \mathbf{v}_{i-r+s} \text{ XOR SR}(\mathbf{v}_{i-r}) \text{ XOR } \mathbf{a}, \end{aligned}$$

where SR is the one-bit right-shift operation.

However, the following result has been found [12]:

Proposition 1. *Any Twisted GFSR in (4) is k -distributed with at most 2-bit resolution for all $k > r$.*

After this finding, Matsumoto and Kurita [7] investigated more general Twisted GFSRs, i.e., $A = PC^T P^{-1}$ in (3), where P is an $L \times L$ nonsingular matrix over F_2 , to obtain the following:

Proposition 2. *Any Twisted GFSR in (3) has an order of equidistribution of at most $r[L/l]$ for the leading l bits.*

Note that the maximum order of equidistribution for GFSR sequences with a period $2^{Lr} - 1$ is $[Lr/l]$ for the leading l bits. In addition, we already have empirical evidence [2, 10] that it is easy to find GFSR sequences having the maximum order of equidistribution for any number of leading bits. For example, comparing $[Lr/l] = 66$ with $r[L/l] = 50$ for $l = 12$, $L = 32$, and $r = 25$, we see that Twisted GFSRs in (3) are not satisfactory compared with general GFSRs.

Thus, the current research on Twisted GFSR generators is to find optimal sequences in the above sense. One idea is to combine several Twisted GFSRs, since Tezuka and L'Ecuyer [13] showed that combinations of Tausworthe sequences are an efficient approach to obtaining sequences with desirable properties in high dimensions. In general, we define the combined GFSR sequence as follows:

$$U_i = u_i^{(1)} \text{ XOR } \dots \text{ XOR } u_i^{(J)},$$

where for $j = 1, \dots, J$, each sequence $u_i^{(j)}$, $i = 1, 2, \dots$, is a GFSR sequence. In this paper, we assume the periods of components are pairwise coprime, so that the period of the combined sequence U_i is equal to the product of all the periods of $u_i^{(j)}$, $j = 1, \dots, J$.

2.2. Couture's Theorem. Following Couture et al. [1], we use the following operators:

$$\begin{aligned} \text{frac}(x) &= \alpha_{-1}z^{-1} + \alpha_{-2}z^{-2} + \dots, \\ \text{trunc}_l(x) &= \alpha_n z^n + \alpha_{n-1}z^{n-1} + \dots + \alpha_{-l}z^{-l}, \end{aligned}$$

where x is an element in the field K of formal Laurent expansions (at infinity) with coefficients in the Galois field F_2 :

$$(5) \quad x = \alpha_n z^n + \alpha_{n-1}z^{n-1} + \dots,$$

where n is any integer and $l \in \mathbb{Z}$.

Define a non-Archimedean valuation in K by

$$|x| = \begin{cases} 0 & \text{if } x = 0, \\ 2^{-n} & \text{if } x \neq 0 \text{ and } x \text{ is given by (5) with } \alpha_n \neq 0. \end{cases}$$

Let k denote a positive integer. The vector space K^k is then normed by $\|X\| = \max_{1 \leq i \leq k} |x_i|$, where $X = (x_1, \dots, x_k)$. Let $A = F_2[z]$ be the subring of polynomials. We call a free A -submodule of K^k a *lattice*.

Definition 3. Let X_1, \dots, X_h be points in a lattice $\mathcal{L} \subset K^k$ of rank h . We call X_1, \dots, X_h a *reduced basis* of \mathcal{L} over A (in the sense of Minkowski) if the following properties hold:

- (i) X_1 is a shortest nonzero vector in \mathcal{L} ;
- (ii) for $i = 2, \dots, h$, the vector X_i is shortest among the set of vectors X in \mathcal{L} such that X_1, \dots, X_{i-1}, X are linearly independent over A .

The numbers $\sigma_i = \|X_i\| > 0$ are then uniquely determined by the lattice, and $s_i = \log_2 \sigma_i$ for $i = 1, \dots, h$ are called its *successive minima*. Lenstra [4] gives details of how to compute these numbers efficiently.

Let $C_r^{(k)} = \{X \mid \|X\| < 2^r\}$, $r \in \mathbb{Z}$. Then, one can view $\mathcal{L} \cap C_r^{(k)}$ as a vector space over F_2 , with cardinality 2^d , where d is its finite dimension over F_2 . The next theorem shows that the number 2^d of lattice points in the cube $C_r^{(k)}$ is determined by r and the lattice's successive minima.

Theorem 1 (Couture). *We have*

$$d = \sum_{i=1}^h (r - s_i)^+,$$

where t^+ denotes $\max(t, 0)$ for a real number t .

For each integer $l \geq 0$, let $E_l^{(k)} = \text{trunc}_l(C_0^{(k)})$, where trunc_l is applied componentwise. Let S be a subset of $C_0^{(k)}$. We now define a frequency function $f_l: E_l^{(k)} \rightarrow N \cup \{0\}$ by

$$f_l(X) = \text{card}\{R \in S \mid \text{trunc}_l(R) = X\}.$$

The set $E_l^{(k)}$ corresponds to a partition of the hypercube $[0, 1)^k$ into 2^{lk} cubic cells of the same size. We note that, if $X \in E_l^{(k)}$ and $R \in S$, the condition $\text{trunc}_l(R) = X$ means that the point in the k -dimensional Euclidean space corresponding to R lies inside the cube $\prod_{i=1}^k [x_i, x_i + 2^{-l})$, where x_i is the real number corresponding to the i th coordinate of X , and $f_l(X)$ is then the number of such points $R \in S$ falling into this cube. For each integer n , let

$$(6) \quad \varphi_{l,k}(n) = \text{card}\{X \in E_l^{(k)} \mid f_l(X) = n\},$$

which represents the number of cells that contain exactly n points. Couture et al. [1] investigated the point set S obtained from linear congruential sequences in K , and developed an efficient method to calculate $\varphi_{l,k}(n)$ by using the above theorem. We will be concerned in the next sections with the problem

of computing $\varphi_{l,k}(n)$ efficiently for the point set S obtained from combined GFSR sequences.

3. MAIN RESULT

First, we give the definition of the resolution-wise lattice of GFSR sequences.

Definition 4. Let a GFSR sequence be given as in (2). Let $g_l(z) = z^{j_l} \pmod{M(z)}$ with $\deg(g_l) < \deg(M)$ for $l = 1, \dots, L$. Then we define the resolution-wise lattice of the sequence as follows: for each $l = 1, 2, \dots, L$,

$$\mathcal{L} = AR_0 + A^l,$$

where

$$R_0 = (g_1/M, \dots, g_L/M).$$

The ordinary lattice can be called the “dimension-wise” lattice, and is useful for the analysis of simple and combined Tausworthe sequences [1]. However, as I have pointed out in [11, 12], the general GFSR sequences cannot be formulated as linear congruential sequences in K . Hence, it is difficult to directly apply Couture’s theorem to the analysis of their k -dimensional distribution. That is why we introduced the notion of a resolution-wise lattice. The following results show the usefulness of this notion.

Proposition 3. For a simple GFSR sequence u_i , $i = 1, 2, \dots$, in (2), define $R_i = \text{frac}(zR_{i-1})$, $i = 1, 2, \dots$, where frac is applied componentwise. Then we have

$$\varphi_{l,k}(n) = \bar{\varphi}_{k,l}(n),$$

where $\varphi_{l,k}(n)$ corresponds to the point set $S = \{(u_i, \dots, u_{i+k-1}), i = 1, 2, \dots\}$, and $\bar{\varphi}_{k,l}(n)$ corresponds to $\bar{S} = \{R_i, i = 1, 2, \dots, |\text{trunc}_k(R_i) = X \text{ for } X \in E_k^{(l)}\}$.

Proof. This follows from the fact that both $\varphi_{l,k}(n)$ and $\bar{\varphi}_{k,l}(n)$ correspond to $S(l, k)$. \square

Since the evolution of R_i , $i = 1, 2, \dots$, gives $2^r - 1$ distinct points in $\mathcal{L} \cap C_0^{(l)}$, we can exploit Couture’s approach to calculate $\bar{\varphi}_{k,l}(n)$, i.e., $\varphi_{l,k}(n)$. We obtain the following result based on Couture et al. [1, Corollary 1].

Proposition 4. A simple GFSR sequence has the order of equidistribution $-s_l$ for the leading l bits, where s_l is the l th successive minimum of the resolution-wise lattice associated with the sequence.

Proof. Switch the interpretation of resolution and dimension in Couture et al. [1, Corollary 1]. \square

Now, we have the main result:

Proposition 5. For a combined GFSR sequence U_i , $i = 0, 1, \dots$, define $\bar{R}_i = R_i^{(1)} \text{ XOR } \dots \text{ XOR } R_i^{(J)}$. Then we have

$$\varphi_{l,k}(n) = \bar{\varphi}_{k,l}(n),$$

5. DISCUSSION

As shown in [12], the original GFSR sequences defined in (1) can be formulated as linear congruential sequences in K . Thus, the k -dimensional distribution of this class of sequences can be analyzed by using the dimension-wise lattice structure as well as the resolution-wise one. Taking this case, we consider the advantage of the approach based on the resolution-wise lattice with respect to practical efficiency. Lenstra [4] shows that his basis reduction algorithm runs in $O(B^2k^4)$, where B is the degree of $M(z)$ and k is the maximum number of dimensions. Therefore, if we use the conventional dimension-wise lattice analysis for high-dimensional behavior of GFSR sequences, the values of k and B are usually 500 or more. On the other hand, in the case of resolution-wise lattice analysis, while B is just as large, k is at most 32 (for a 32-bit computer). Obviously, the latter approach is practically better.

ACKNOWLEDGMENT

The author is grateful to Raymond Couture, Pierre L'Ecuyer, and Makoto Matsumoto for their comments.

6. APPENDIX: SUCCESSIVE MINIMA FOR THE TWO TWISTED GFSRS AND THEIR COMBINATION

TABLE 2. The first successive minimum $-s_1$ of the resolution-wise lattice in resolutions 2 to 31 for the Twisted GFSR with a period of $2^{403} - 1$. All other successive minima s_2, \dots, s_l are equal to -13 for all resolutions

res.	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$-s_1$	390	377	364	351	338	325	312	299	286	273	260	247	234	221	208
res.	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
$-s_1$	195	182	169	156	143	130	117	104	91	78	65	52	39	26	13

TABLE 3. The first successive minimum $-s_1$ of the resolution-wise lattice in resolutions 2 to 31 for the Twisted GFSR with a period of $2^{800} - 1$. All other successive minima s_2, \dots, s_l are equal to -25 for all resolutions

res.	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$-s_1$	775	750	725	700	675	650	625	600	575	550	525	500	475	450	425
res.	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
$-s_1$	400	375	350	325	300	275	250	225	200	175	150	125	100	75	50

TABLE 4. The successive minima s_1, \dots, s_4 of the resolution-wise lattice in resolutions 2 to 31 for the combined Twisted GFSR with a period of $(2^{403} - 1)(2^{800} - 1)$. All other successive minima s_5, \dots, s_l are equal to -38 for all resolutions $l = 5, \dots, 31$

res.	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$-s_1$	775	750	725	700	675	650	625	600	575	550	525	500	475	450	425
$-s_2$	428	380	367	354	341	328	315	302	289	276	263	250	237	224	211
$-s_3$	-	73	60	60	60	60	60	60	60	60	60	60	60	60	60
$-s_4$	-	-	51	51	51	51	51	51	51	51	51	51	51	51	51
res.	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
$-s_1$	400	375	350	325	300	275	250	225	200	175	150	125	100	75	50
$-s_2$	198	185	172	159	146	133	120	107	94	81	68	58	51	47	44
$-s_3$	60	60	60	60	60	60	60	60	60	60	60	57	51	47	42
$-s_4$	51	51	51	51	51	51	51	51	51	51	51	51	51	46	41

BIBLIOGRAPHY

1. R. Couture, P. L'Ecuyer, and S. Tezuka, *On the distribution of k -dimensional vectors for simple and combined Tausworthe sequences*, Math. Comp. **60** (1993), 749–761.
2. R. F. Koopman, *The order of equidistribution of subsequences of some asymptotically random sequences*, Comm. ACM **29** (1986), 802–806.
3. P. L'Ecuyer, *Uniform random number generation*, Ann. Oper. Res. (1994) (to appear).
4. A. K. Lenstra, *Factoring multivariate polynomials over finite fields*, J. Comput. System. Sci. **30** (1985), 235–248.
5. T. G. Lewis and W. H. Payne, *Generalized feedback shift register pseudorandom number algorithms*, J. Assoc. Comput. Mach. **20** (1973), 456–468.
6. M. Matsumoto and Y. Kurita, *Twisted GFSR generators*, ACM Trans. Modeling and Computer Simulation **2** (1992), 179–194.
7. ———, *Well tempered twisted GFSR generators*, manuscript, (1992).
8. H. Niederreiter, *Random number generation and quasi-Monte Carlo methods*, CBMS-NSF Regional Conference Series in Applied Math., no. 63, SIAM, Philadelphia, PA, 1992.
9. S. Tezuka, *Walsh-spectral test for GFSR pseudorandom number generators*, Comm. ACM **30** (1987), 731–735.
10. ———, *A heuristic approach for finding asymptotically random GFSR generators*, J. Inform. Process. **10** (1987), 178–182.
11. ———, *Lattice structure of pseudorandom sequences from shift register generators*, Proc. 1990 Winter Simulation Conference, IEEE Press, 1990, pp. 266–269.
12. ———, *A unified view of long period random number generators*, submitted, (1992).
13. S. Tezuka and P. L'Ecuyer, *Efficient and portable combined Tausworthe random number generators*, ACM Trans. Modeling and Computer Simulation **1** (1991), 99–112.

IBM RESEARCH, TOKYO RESEARCH LABORATORY, 1623-14 SHIMOTSURUMA, YAMATO-SHI,
KANAGAWA 242, JAPAN
E-mail address: tezuka@trlvm.vnet.ibm.com